PATENT ABSTRACTS OF JAPAN

(11)Publication number:

2001-034619

(43)Date of publication of application: 09.02.2001

(51)Int.CI.

G06F 17/30

(21)Application number: 11-203908

(71)Applicant: FUJITSU LTD

(22)Date of filing:

16.07.1999

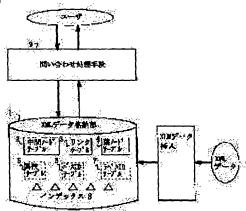
(72)Inventor: KANEMASA YASUHIKO

KUBOTA KAZUMI ISHIKAWA HIROSHI

(54) STORE AND RETRIEVAL METHOD OF XML DATA, AND XML DATA RETRIEVAL SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To make storable XML data into a data base and to make executable a complicated inquiry at a high speed. SOLUTION: A relation data base of an XML data store means 1 includes an intermediate node table 2 which stores the intermediate node information, a link table 3 which stores the link information, a leaf node table 4 which stores the leaf nodes, an attribute table 5 which stores the attribute information, a path ID table 6 where the path IDs are made to correspond to the character strings and a label ID table 7 where the label Ids are made to correspond to the character strings. The XML data which are expressed in a tree structure are divided into nodes, and these nodes are made to correspond to the link information and stored in the tables 2–7. When the XML data are retrieved, an inquiry statement is given to an inquiry processing means 9. The means 9 executes an inquiry to track a tree structure by using index 8 and outputs a requested retrieval result.



LEGAL STATUS

[Date of request for examination]

06.09.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本國特許庁 (JP)

(12)公開特許公報(A) (11)特辦出屬公開番号

传開2001-34619 (P2001~34619A) (3)公開日 平成13年2月9日(2001.2.9)

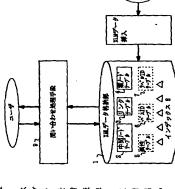
| | | | 23 | 3)公開日 | (P2001-34619A) (43)公明日 平成13年2月9日(2001.2.9 |
|------------------|------|---------|--------|-------|---|
| (51) Int. C.I. 7 | 微別記号 | r. | | | デーマコード (参考) |
| G 0 6 F 17/30 | | G 0 6 F | 15/419 | 320 | 58075 |
| | | | 15/403 | 330 | 13 |
| | | | | 3 4 0 | ۵ |

| | 独強請求 非請求 請求項の数5 | OL | (全15頁) |
|-----------|-----------------------|-----------------------------|--------------------------------------|
| (21) 出版番号 | 特 例平11-203908 | (71) 出版人 | (7]) 出題人 000005223 |
| (22) 出版日 | 平成11年7月16日(1999.7.16) | | 鱼工遗体式受在 神奈川県川崎市中原区上小田中4丁目1番1 导 |
| | | (72) 発明者 | 金政 泰彦 神奈川県川崎市中原区上小田中4丁目1番1 |
| | | (72) 発明者 | 号 富士通株式会社内 从保田 和己 |
| | | | 神奈川県川崎市中原区上小田中4丁目1番1号 - 富士道株式会社内 |
| | | (74) 代理人 100100930 井理士 長 | 100100930 弁理士 長臀 俊一郎 (外1名) |
| | | | 放送回に扱く |

[54] 【発明の名称】XMLデータの格納/依案方法およびXMLデータ依案システム

本発用の基本構成図

(57) [契約] 【課題】 XMLデータをデータベースに格納し、複雑



[特許請求の範囲]

【精米項1】 XMLで配述されたデータを、エレメントを中国ノードとし、エレメント値と原件値を竣ノードとし、エクストト値と原件値を竣ノードとし、タグをリンクとする木梯通で表現し、

NNLの木構造をノードとリングに分解し、各ノードとリンク指数を収録されているカードとリングに分解し、各ノードとリンク指数を観響を開発されて図像データスースのテーブルに

上記関係データベースに格納されたデーブルを利用して、アカチルギモンに、カースに格納されたデーブルを利用して、アカーギージャージャを発

て、任意の格道のXMLデータを依案することを特徴とするXMLデータの格制/検索方法。 「翻求項2」 エレメントを中間ノードとし、エレメント値と属性値を築ノードとし、タグをリンクとする米橋造で表現フードとし、タグをリンクとする米橋造で表現されるXMLで記述されたデータを検索するシ

ステムであって、

上記システムは、XMLデータを格納する格納手段を値え、数格約手段の図条データベースに、少なくとも中間ノードの複数を格納するための中間ノードテーブルと、リンクの情報を格納するためのリンクテーブルと、楽ノーの信報を格納するためのリンクテーブルと、楽ノードの情報を格納するためのジノードテーブルとを数

上記XM上の木精造をノードとリンクに分解して、上記テーブルに各ノードとリンク情報を関係付けて格制し、上記テーブルを参照して木構造を辿る間い合わせを実行し、XM上データを検索することを特徴とするXMLデータを検索システム。

「謝求項3」 関係データベースに、パスの文字列とバス角の10の対応数であるパス10テーブルと、ラベルの文字列とラベル用10の対応数であるラベル10デーブルとを設けたことを特徴とする翻求項2のXMLデータ数案システム。

「値永項4】 リンクテーブルの中に各子エレメントがそのエレメント内で出現した順序の情報を付加し、築ノードテーブルの中に各エレメント値がそのエレメント内で出現した順序の情報を付加し、上記値報により元のXML文庫の復元を可能としたことを特徴とする翻求項2のXMLデータ検索システム。

「翻氷項5」 中国ノードテーブルに、ノードIDによる校業を高速に行なっためのインデックスと、テーブルの文庫IDによる校業を直通に行なっためのインデックスを、パスIDによる校業を直通に行なっためのインデスと、パスIDによる校業を直通に行なっためのインデスと、パスIDによる校業を直通に行なっためのインデ

ックスを用磨し、 リンクテーブルに、親ノードから子ノードを南辺に微珠するためのインデックスと、テノードから殺ノードや酒 踏成数字るためのインデックスと、デノードから殺ノードを酒 語に彼蝶するためのインデックスを用意し、

ボノードデーブルに、ノードIDからそのノードの値を得るためのインデックスと、ある値を持つノードを徴発するためのインデックスを用着し、

と、格納スキーマの数計が問題となる。例えば、図8にを依案するためのインデックスを用意し、 のみば、図8にラベルIDテーブルに、ラベルの文字列に対応するラベ 50 示される 【DTD】を待つ、サンプルXMLデータ【X

パスIDテーブルに、パスの文字列に対応するパスID

8

特別2001-34619

ルーロを検ばするためのインデックスを用着し、 上記インデックスを用いて木帯道を辿る間い合わせを致 行することを特徴とする間状項2、3または翻桨項4の XMLデータ検索システム。

(発明の詳細な説明) [0001] 「発明の図する技術分野」本発明は、XMLで記述された大幅のデータを関係データペースに格勢し、依楽する XMLデータの格耐く破水方法もよび破楽ンスナムに図 D、特に、XML文盤の構造に依存せずにあらゆるXM レータを格割できるようにし、また格輪されたXML データに対するXMLの本機癌を辿る間い合わせを高型 に実行できるようにし、XMLデータの格料人破落方法 および依頼システムに関するものである。 【従来の技術】現在、XMLデータを格納するのに用いられている手柱は、大まかに次の2つのタイプに分類することができる。

[0002]

●ファイル格納:XML文型をファイル形式のまま格約 かる手法。この手柱は、オリジナルのXMLファイルの 会体あるいは一部をそのまま利用することを目的として おり、そのため、XML文報をファイル形式のま指格料 する。しかし、それだけでは、ファイルの数が値えたと きに目的とするファイルを見つけ出すことが困難になる ので、目的とするファイルを扱つけ出すことが困難になる ので、目的とするファイルを発しけますことが困難になる ので、目的とするファイルを発力はあるのインデックス も用着しておく必要がある。 (0003)のテーブル格的:XMLを関係データベースのテーブルにマッピングして格勢する手法。この手指ではXMLX動を指摘化プラと見なし、データベースに格替することによって高温な校表を行なうことを目的 ほしている。そのため、この手指では、各エレメントを開発すータベースのテーブルの各カラムにマッピングして格等する、XMLデータをテーブルにマッピングする おには、XMLの各エレスントをテーブルの名カラムにどのようにマッピングするかには、XMLの各エレスントをテーブルの名カラムにどのようにマッピングするかというマッピングは関係が要である。このマッピング規則はユーザが単値に指定する必要がある。

[0004] [発明が解決しようとする観題] XMLデータを格納する際に一番問題によるのは、そのデータ構造が一葉に流 まっていないという事である。特に、DTD (文據型質 同のないXMLデータでは、どこにどのようなタグが 出致するか分からず、データ構造は全く分からない。D TDのあるXMLデータできえも、DTDの中でタグの 報り返しゃタグの選択、タグの再補的な宣音が構きれて いるので、データ構造が一路に定まらない。なお、この ようなデータを準構造データと呼ぶ。このようなデータ 構造の定まっていないXMLデータを格納しようとする と、格納スキーマの設計が問題となる。例えば、図8に と 格教スキーマの設計が問題となる。例えば、図8に

[0005] 図9は上記XMLデータをテーブルに格納 した以子を示す図である。図9のテーブルでは、1タブ されているタグも格納できない。さらに、そもそもXM レデータにDTDが存在しなくて、どのようなタグが出 ルが本1 冊分の情報に相当していて、列にはXMLデー る。これを見ると、一見サンプルデータが問題なく格絶 されているかのように見える。しかし、サンブルデータ 図9のテーブルでは落者を格納するスペースは最大2人 分しか用意されていない。もしXMLデータの中に蓄者 テーブル格納では、XMLのDTDで配近される線 り返しタグを格納することができない。これは、テーブ ル格紙ではあらかじめ格納する要果を列として指定して おく必要があるので、最大数が未定の織り返し要素を表 現できないからである。また、同じ型由で再幅的に定義 現するか分かっていないときには、テーブルの杮造を決 のDTDに掛かれた定義には著者数の制限が無いのに、 タ中で出現する可能性のある金てのタグがとられてい がそれ以上存任したら、そのデータは格納できないか、 格納しても情報が一部欠損することになる。このよう められず、全く対応できない。

\$ クスが必要となる。インデックスの構成は目的に応じて [0006] 一方、ファイル格納は、XMLデータをフ ァイル形式のまま格殻するので、DTDの無いXMLデ **ータであろうと半格造のXMLデータであろうと、格納** では大国に格納されたデータの中から自分の求める情報 色々と考えられ、簡単なものではタグ名と文字列の組を できないXMLデータは存在しない。 しかし、それだけ だけを⋳穽することができないので、破穽用のインデッ キーにして、そのタグに狙まれてその文字列が出現して る。しかし、そのような簡単なインデックスでは、タグ の階層構造を考慮した檢案は行なえない。タグの階層構 造の情報を持つようにインデックスを工夫することも考 の情報を持っていないので、XMLデータの全情報を使 【0007】〇 インデックスがXMLの木塔造の全て えられるが、それでもなお次のことが問題として残る。 いるようなXML文格を検索してくるというものがあ した政権がいかない。

の インデックスが木構造を辿ることに最適化されていないので、そのような検索を行なった場合は検索選択が

以上のように、データ構造が一般に定まっていないXM レデータにおいては、いかにしてDTD熊しのXMLデータや半構造のXMLデータを格納するか、また、格納されたXMLデータを格納するか、また、格納されたXMLデータに対していかにして未締造を辿るような複雑な問い合わせを偽選に契行できるようにするかといった問題がある。本発明は上記した事情に鑑みなされたものであって、本発明の目的は、データ構造が一巻

2

に定まっていないXMLデータをデータベースに格制し、複雑な間合わせを高遠に気行することができるXMLデータの格制/微築方法およびXMLデータ検索システムを提供することである。

【麒題を解決するための手段】図1は本発明の基本構成 を示す図である。同図に示すように、本発明のシステム は、エレメントを中回ノードとし、エレメント値と阪性 値を取ノードとし、タグをリンクとする木構造で表現さ れるXMLで記述されたデータを検索するシステムにお いて、XMLデータを格納する格納手段1を改け、該格 納手段1の関係データベースに、少なくとも中間ノード の情報を格納するための中間ノードテーブル2と、リン クの情報を格納するためのリンクテーブル3と、縦ノー ドの情報を格納するための菜ノードテーブル4とを散け る。そして、上記XMLの木構造で表現されたXMLデ 木構造を形成する中間ノードと、エレメントの値を持っ ている菜ノードとでは、格納するために最適な格納構造 が異なるので、上記のようにそれぞれ殷適化された別々 値を持つためのノードである珱ノードと木構造の情報を 持っためのノードである中間ノードを別々のテーブルに 格納することにより、値を格納するための格納スペース を節約することが可能となる。各ノード間の接続情報を 保持する為のリンクも、リンクテーブル3に格納して持 っておく必要がある。また、風性情報を格納するための ドテーブル2に各ノードのルートからのフルパス情報を ータをノード単位で分割し、上記テーブル2~4に各ノ **属性テーブル5を別途散けてもよい。さらに、中間ノー** 1 Dで記述し、パス用の1 Dと文字列の対応装をパス1 Dテーブル6として別に持つことにより、格納スペース リンクテーブル3のタグ名と属性ノードテーブルの属性 名を1Dで記述し、これらラベルの1Dと文字列の対応 て、格納スペースの節約と文字列段森の高遠化を図るこ とができる。また、リンクテーブル3の中に各子エレメ の専用テーブルに格納するのが望ましい。このように、 ードとリンク情報を関係付けて格納する。XMLでは、 の節約と、校案の萬速化を図ることができる。同様に、 液をラベル 1 ロテーブル7 として別に持つことによっ ントがそのエレメント内で出現した順序の情報を付加 22 20

【0009】本発明では、XMLの本構造をそのまま格 毎年段1に格納するので、DTD無しのXMLデータキ 半借道のXMLデータも格納できる。また、XMLの木 構造を全てデータベース上に格約しているので、本構造 をなてび解報を検察に利用することができる。しかして かだけでは同い合わせが行なわれたときに、ノード単位 に分割して格納されているXMLデータの本構造を再結 合するのに時間がかかり、関い合わせの実行時間が通く

なる。そこで本発明では、上記のテーブル2~7に、X MLデータへの問い合わせパターンを考慮してインデックス8を張る。これにより、XMLの本格強を辿るような機能な問い合わせの実行を高速に行なうことを可能となる。上記XMLデータを検索するには、例えばXMLデータを検索するには、例えばXMLデータを検索するには、例えばXMLデータも性の目い合わせのための構文を全止成し、環路を検察の存行のはかられて表が多くに、本格造検照の関数セットで記述される。この実行プランに、本格造検照のの数を上りて表れて表して、要求された検索をして本格強を起る問い合わせを実行し、要求された検索を出いて本格強を起る問い合わせを実行し、要求された検索を担めする。

[0010] 本発列においては、次のように構成することもできる。

(1) テーブルに関係データベースの顧約の機能を適用することによって、XMLの構文規則をチェックする。(2) リングテーブルの中に、各エレメントの同ラベルを持つ記却エレメント中での出現順序の情報を付加し、各ラベルの出現順序を指定した問い合わせの実行を可能

(3) リンクテーブルにリンクの両部点の捕船だけでなくタグ名の揺棄も待つことによって、タグ名を指定してリンクを辿る回い合わせを高速に実行する。

(4) 風性テーブルの中の風性ノードの接続先をリンクではなくて中間ノードこすることによって、風性を条件にして木構造を辿る間い合わせを実行する際のテーブル砂楽回数を削減し、関い合わせの高速実行を可能とす

(5)中間ノードテーブルのバスIDによる役権を直避 に行なうためのインデックスをB・-1ce で海鎖する場合において、キー値をバスIDとノードIDの箱とすることによってキー値の風視を無くす。

(6)中間ノードテーブルの文章1Dによる役余を高遠に行なうためのインデックスをB・-1ree で格察する場合において、キー値を文章1Dとノード1Dの総とすることによってキー値の監視を無くす。

[0011]

【発明の実施の形態】以下、本発明の実施の形態について説明する。

(1) システム構成

し、採ノードテーブルの中に各エレメント値がそのエレ

メント内で出現した順序の情報を付加することにより、

元のXML文替の復元が可能となる。

図2は本発明の実施例のシステムの構成を示す図である。 同図に示すように、本実施例のシステムは大きくわけて、XMLデータ格解部11、XMLデータ格解部11にXMLデータイを対して、XMLデータイを対して、AMを対して、AMLデータインにもなる。2muをのではかけないないが、XMLデータインを行う。によって、XMLデータ格解部11に挿入される。XMLデータインを対しま、XMLバーが12aとローダー12bから成り、XMLバーザ12aとコーダー12bから成り、XMLバーザ12aとコーダー12bから成り、XMLバーザ12aとコーダー12bから成り、XMLバーザ12aは

TU3.

特開2001-34619

3

れたXMLデータを指文解析し、XMLデータの末構造を、XMLデータ格料部11に格納できるようにノード単位に分解する。また、ローダー12 bは、そのノード単位に分解された木構造をXMLデータ格制部11のテーブルに高コまえ

において、XMLファイルを競み込む。ステップS2に ノード情報とリンク情報を中間形式としてファイル出力 する。また、解析が成功しない場合には、構文解析失敗 としてエラー出力し処理を終了する。ステップS4にお いて、生成された中間形式ファイルを視み込み、ステッ って関係データベースの各テーブルに何入し、処理を終 [0012] 図3に上紀XMLデータの格徴処理を示す フローチャートを示す。本変施例においてXMLデータ の格納処理は次のように行われる。まず、ステップ51 おいて、XMLパーザにより、入力ファイルの結文解析 き、XMLバーザが解析結果として、XMLの木構造の プS 5において、飲み込んだXMLデータをローダによ 了する。また、上配値入が成功しない場合には、データ を行う。解析が成功した場合には、ステップS3に行 **ゆ入失敗としてエラー出力をして処理を終了する。** 2 20

[0013] 格納されたXMLデータに対する問い合わ せは、XMLデータ問い合わせ宮稲で行なわれ、その間 問い合わせ処理エンジン13は、何い合わせ言語のパー ザ13a、聞い合わせ駁踏化エンジン13b、木帯造板 13aは、入力された問い合わせ文の構文チェックを行 い問い合わせのための構文木を生成する。問い合わせ最 適化エンジン13bは、上紀構文木を基に、殷適な実行 プランを生成する。この実行プランは、木構造検索用A P 1 1 3 c は、XMLデータ格触部 1 1 とのインタフェ ースで、XMLの木構造上での基本的な検索を行なう関 **密用API (アプリケーション・プログラミング・イン** タフェース)13cから成る。問い合わせ宮部のハーザ PI13cの関数セットで記述される。木構造検索用A い合わせは問い合わせ処理エンジン13で処理される。 数のセットである。 30

[0014]次に、上紀システムにおける各部の構成に ついてきらに詳細に説明する。

(1) テーブル構成

まず、上記XMLデータ格軸部11に格軸されるテープ40 ルの海底について説明する。XMLデータを本帯造で設現する力法はいくつかあるが、本政施例では図4に示す本地造設現を憩定している。図4は、前記図8に示しXXMLデータを本帯造で扱現したものである。この本街造投現において、丸い中間ノードはエレメントを装しており、ノードの観子図係がエレメントの包含図係を表し

(0015)また、ノードの丸の中の数字はノードID を投している。ノードとノードを結ぶリンク(枝)はケグを投しており、リンクの荷に築かれている文字列はグ50 グ名を投している。三角の葉ノードはエレメントの値を

特閒2001-34619

クの情報が欠落してしまう。そこで、リンクの情報はリ ノード、頃性の双ノードとは吸適な格納構造が異なるの 数し、四角い菜ノードはタグに付けられた風性 Mirrib uic)を表している。値を持つのはこの2つの菜ノードだ けである。ノードを分割してデータベースに格納すると きに、ノードの借報だけをデータベースのテーブルに格 始したのでは、木紙道のノード間の繋がり、つまりリン ンクの情報としてそれを格納する専用のテーブルを用意 する。またノードも、中間ノードと、エレメント値の漿 で、別々のテーブルに格納する必要がある。

【0016】本実施例で使用するテーブルは、全部で次

○中国ノードテーブル の6つである.

これは中間ノードの情報を格納するテーブルである。ノ ードID (id)の他に、そのノードが合まれている文協の 文偽ID (docid) 、 そのノードまでのルートからのフル パスのID (pathid) をカラムとして持っている。 ゆリンクテーブル

ード1 D (id)、リンクのラベル (タグ名) の1 D (labe 全兄弟ノード中での出現順序(tord:total order)、その lid)、子ノードのノードID(child)、その子ノードの ナノードの何ラベルを持つ兄弟ノード中での出現順序も ord:partial order)をカラムとして持っている。上記の ように、リンクテーブル中にラベル (タグ名) の10() これはノード信のリンクを格納するテーブルである。ノ a belid) を付加することによりタグ名を指定してリンク を辿る問い合わせを高遠に実行することが可能となる。 [0017] 母派ノードテーブル

る。そのエレメントにあたる中間ノードのノード1D(i これはエレメント値の張ノードを格納するテーブルであ d)の色に、エフメントの魚(value) と、そのエフメント 中でその値が出現した順序(order) をカラムとして持っ ている。このように、値を持つための菜ノードテーブル を、前記中間ノードテーブルとは別に散けることによ り、値を格納するスペースを節約することができる。 [0018] ④成位ノードテーブル

\$ のノード I D (id)の包に、図在名の I D (labelid) 、図 これはタグにつけられた原性(例えば図8における合000 k year="1995">におけるyear) を格納するテーブルであ る。そのタグが含まれるエレメントにあたる中間ノード 作値 (Alivalue) をカラムとして待つ。なお、風性テープ ルに関係データベースの調約機能を用いて、(id, labeli て、「同一のタグ内では同一の风性名は出現してはなら ない」というXMLの原性に関する構文規則をチェック することができる。また、本実施例で想定している木橋 過表現では、XMLのタグが木構造のリンクに相当する ので、XMLのタグに付けられる原作は本来ならばリン クに付くべきである。しかし、図4では、図性はリンク むの組がユニークという 固粒をかけて おくことによっ

る。すなわち、原性を条件として木構造を辿る関い合わ は、検索時のテーブル参照の回数を少なくするためであ せを実行する際のテーブル検索回数を削減し、問い合わ せの高速化を図ることが可能となる。

[0019] あパス! ロテーブル

これはパスIDとパスの文字列の対応表である。パスの 文字列を中間ノードテーブルに直接格き込まないでこの ように別に持っているのは、スペースの節約の為もある が、パス名の文字列マッチングを含む検索が行なわれた ときに、検索対象が少なくてすみ、検索が高速化できる

あラベル I Dテーブル

からでもある。

ブルの腐性名を1Dで記述し、このラベルの1Dと文字 **列の対応数をラベル I Dテーブルとして別に持つことに** のように、リンクテーブルのタグ名と、風性ノードテ より、パスIDテーブルと同様、格納スペースの節約 これはラベルIDとラベルの文字列の対応表である。 と、故案の南遠化を図ることができる。

に、子ノードの全兄弟ノード中での出現順序(lord:lola に、各エレメント値がそのエレメント内で出現した順序 (order) の債報を付加することに、XMLデータ格納部 から、元のXML文書を復元することが可能となる。例 えば、「今日は〈天気〉晴れく/天気〉だった。〇〇は 〈 協所〉デバート(人場所)へでかけた。」のようにタグで ンクテーブル中に、各エレメントの同ラベルを持つ兄弟 ノード中での出現駅序 (pord:parlial order)の情報を付 1.1に格納されるノード単位に分解されたXMLデータ 区切られた文章を復元することも可能になる。また、リ 加することにより、各ラベルの出現順序を指定した問い lorder)の情報を付加し、また、菜ノードテーブル中 [0020]また、上記のように、リンクテーブル中 合わせを高速に実行することが可能となる。 20 30

[0021] - 宮として、図8のサンブルXMLデータ (図4の木構造表現)を上記のテーブル群で格納した様 子を図5、図6に示す。図5は中間ノードテーブル、リ ンクテーブルの例を示す図である。中間ノードテーブル において、例えば、第1行目のid (=5) は図4におい て"5"と記されたノードを示し、そのノードが含まれ ている文樹の文替 I D (docid) は1である。また、その ノードまでのルートからのフルバスのID (pathid) は1 であり、この I Dに対応したpathは、"bib. book publi sher. name"である。また、リンクテーブルにおいて、例 えば1行目のid (=4) は図4において、"4"と記さ れたノードを示し、そのJabelidは5であり、このJabe id に対応するlabel は"name"である。また、その出現 順序を示す(ord, pord はそれぞれ" 0"," 0" であり、 子ノードは、図4で"5"と記されたノードである。

ブル、パスIDテーブル、ラベルIDテーブルの例を示 す図である。 菓ノードテーブルにおいて、例えば第1行 [0022] 図6は漿ノードテーブル、風性ノードテー

2

に対してではなく、その下のノードに付いている。これ

れ、例えば、palhid=" 1" に対応した文字列は前記し ルにおいて、例えば第1行目のid (=3) は図4におい 95"である。また、パスIDテーブル、ラベルIDテ ドを示し、そのorder は"0"、またその菜ノードの値 (value)は"Addison-Wesley"である。 風性ノードテーブ て、" 3" と記されたノードを示し、そのlabelid は3 ("year"に対応), その原性値 (ailvalue) は"19 ーブルにはそれぞれ、上記各テーブル中のpathid、labe lid に対応したパスの文字列、ラベルの文字列が格納さ 例えばlabelid =" 1" に対応した文字列は" bib"であ **目のid (=5) は図4において、"5" と記されたノー** たように" bib. book. bn blisher. name"であり、また、

作が行なわれる。このジョイン操作の速度は全体の検索 本実施例においては、本来連結されていたはずの木構造 のノードが、前記したように1つ1つに分割されて関係 に、木構造を辿る問い合わせが行なわれた場合、問い合 わせで辿る部分のリンクを連結し直すためにジョイン機 協強に大むへ影撃するのか、ジョイン豪作や夷湖に行な えるようにインデックスを効果的に張っておく必要があ る。また、問い合わせが行なわれる場合、検索条件とし て指定されるのは、エレメントの値、原性、パス、出現 順序などである。それらの検索も高速に行なう必要があ **あので、そこにもインデックスを用強しておく必要があ** データベースのテーブルに格納されている。このため [0023] (2) インデックスの結成

は Bt -(ree で扱ってあり、キーが複数の既性の組から [0024] 図1に、上記図5、図6に示したテーブル に扱ったインデックスの一覧を示す。 いのインデックス なるインデックスは、その組の先頭からの部分的な原性 ブルに張ってあるインデックスでキーが (pathid, id) の は、一見pathid単独で構わないように思われるかもしれ ない。しかしキーをpathidだけにすると、同じキー値を スが機能しなくなる。上記のようにキー値をパスIDゆ athid) とノードの1口 (id) の組とすることにより、キー 個の風複を無くすいとができ、B → −lrec の核紫や超湿 に行うことができる。また、中間ノードテーブルに扱っ の組で検索に用いることもできる。なお中間ノードテー ものは、あるパスに該当する全てのノードを検索してく 持つエントリが多位に発生して、B * -1ree インデック より、キー値の監復を無くすことができ、B ⁺ -(ree の るときに使用するものである。このインデックスのキー 文芸ID (docid) とノードのID (id)の組とすることに てあるインデックスでキーが(docid, id) も同様であり、 **検索を高速に行うことができる。**

われる。XMLデータのための検案官語の一つとして検 前記したように、格納されたXMLデータに対する問い 合わせは、例えばXMLデータの問い合わせ言紹で行な [0025] (3) 関い合わせの実行

特開2001-34619

9

案官語XQLがある。XQLによる問い合わせ文を、例

SELECT result: <\$book title>

WHERE \$book, author, lastname="Darwen"; FROM book: bib. book

この問い合わせの意味は「bib book, author. lastnameが Darwenであるようなbib. bookについて、bib. book. title を検済結果として得たい」という意味である。

森の条件のセレクションを指定する。上記のような問い で処理される。 聞い合わせ処理エンジン13では、上記 **愚適な実行プランを生成する。この実行プランは、木梢** SELECTの部分では检索結果として得たいエレメントのプ ロジェクションを指定する。FROMの部分では検索の対象 となるエレメントを指定している。WECE の部分では検 合わせは前記したように、問い合わせ処理エンジン13 のような問い合わせ文の樹文チェックを行い問い合わせ <. SELECT、FROM、WHERE の3つの部分に別れている。 のための構文木を生成する。そして、該構文木を基に、 [0027] 上記に示すように、問い合わせ文は大き 20

は、図8のサンプルXMLデータを、XMLデータ格粧 **部11に格制し、自送した図5、図6に示したテーブル** [0028]次に、上紀XMLデータに対する順い合わ せ処理が、どのように行なわれるかを説明する。ここで に挿入した場合を倒として、上記のように「著者がOarw enである本のタイトルを求めよ」という問い合わせを行 は、次のように行われる。なお、下記1. ~10. の処 なった場合について説明する。この場合のテーブル検索 理は、上記木構造検済用の関数により実行される。 造検容用の関数セットで配述される。

[0029] 1. 苹ノードテーブルを検索して、値が **'Darwen'' であるノードのノードID (=16) を得** 2, パスIDテーブルを検索して、パス"bib. book. auth 3. 中間ノードテーブルを上記1. で得られたノード1 or, lasiname " のパスID (=4) を得る。

D (=16) で夜垛して、飾られたパス1D (=4) が 上記2. で得られたパス1D (=4) と一致することを

4. ラベルIDテーブルを投発して、ラベル"lasiname のラベル10 (=8) を待る。 40

5. リンクテーブルを位落して、上記1. で得られたノ **一ドID (=16) と上記4. で得られたラベルID** (=8) から、親ノードのノードID (=15) を得

6. ラベルIDテーブルを拉案して、ラベル"author のラベル10 (=7)を得る。

7. リンクテーブルを傚楽して、上記5. で得られたノ (=1) から、親ノードのノードID (=9)を得る。 **一ドID (=15) と上記6. で得られたラベルID** 22 特開2001-34619

8

8. ラベル1 ロテーブルを位案して、ラベル"[ille"の ラベル10 (=6)を得る。

9. リンクテーブルを検索して上記?. で得られたノー ド1D (=9) と上記8. で得られたラベルID (= 6) から、子ノードのノード1D (=12) を得る。

10. 菜ノードテーブルを検索して、上記9. で得られ たノードID (=12)から、そのノードの銜("Foundal のようにして得られた検索結果は、問い合わせ処理エン ion for Object/Relational Database") を得る。以上 ジン13を介して出力され、ユーザに提示される。

[発明の効果] 以上説明したように、本発明において [0030]

は、関係データベースに、中間ノードの情報を格納する ための中間ノードテーブルと、リンクの情報を格納する めの埰ノードテーブル等のテーブルを散け、XMLの木 ためのリンクテーブルと、災ノードの情報を格納するた 構造をノードとリンクに分解して、上記テーブルに各ノ

夕を検案するようにしたので、データ構造が一意に定ま ードとリンク情報を図係付けて格納し、上記テーブルを 参照して木格造を辿る周い合わせを実行し、ΧMLデー っていないXMLデータに対する複雑な問い合わせを高

遊に実行することができる。また、XMLの木構造をそ のまま格納手段に格納するので、DTD無しのXMLデ

いるので、木樹造の全ての僣粗を検索に利用することが さらにXMLの木樹造を全てデータベース上に格納して ータや半構造のXMLデータも格納することができる。

[図前の簡単な説明]

[図1] 本発明の基本構成図である。

[図2] 本発明の実施例のシステムの構成例を示す図で 30 13 b

特閥2001-34619

[図3] 本発明の実施例のシステムにおける格納処理フ ローを示す図である。

【図4】 XMLデータの木構造表現の一例を示す図であ

【図5】 本発明の実施例のテーブル構成の一例を示す図 (1) である。

[図6] 本発明の実施例のテーブル構成の一例を示す図 (2) である。 [図7] 本発明の実施例のイッデックス一覧を示す図で

848

2

[図9] 図8のXMLデータをテーブルに格納した様子 [図8] XMLデータの一例を示す図である。 を示す図である。

XMLデータ格納格納手段 [符号の説明]

中間ノードテーブル

リンクテーブル

採ノードテーブル

成在ナーブル

バスIDテーブル

ラベルIDテーブル

問い合わせ処理手段 インデックス

XMLデータ格納部

XMLデータ挿入モジュール

XMLバーザ 12a

問い合わせ処理エンジン部 1 2 b

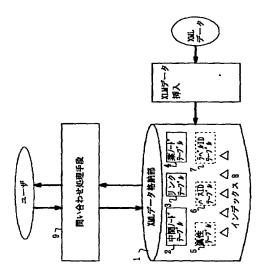
問い合わせ最適化エンジン 問い合わせ言語のパーザ 1 3 a

木桥造検索用

⊛

[図]

本発明の基本構成図



[図7]

本発明の実施例のイッデックス一覧を示す図

インアングスー製

| | | | _ | | | | | | | | |
|--------|-------|-------------|---------------|----------------------|-----------------------|------|-------|-------|----------|-------|-------|
| | P\$ | (docid, id) | (pathid, id) | (id, labelid, child) | (child, labelid, id) | PI | value | 16 | attvalue | path | label |
| 7-711A | 中間ノード | 中間ノード | オーへ図中 | 120 | 11.0 | 業ノード | 業ノード | 異性ノード | オー/ 尹笙 | AZ ID | 54VID |

[图3]

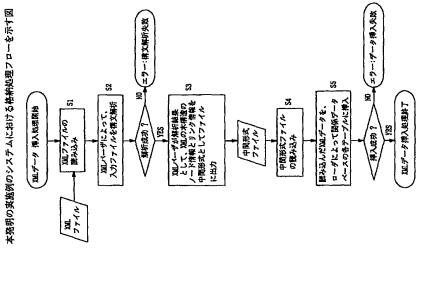
倒い合わせ名組ェンジン13

向い合わせ最適化インジン

木構造檢案用API

困い合わせ言語のパーザ

JULデータ問い合わせ言語



[図4]

"Darweri"

ementesi

.1868.

"Date"

"Foundation for Object/Relational Database"

本発明の実施例のテーブル構成の一倒を示す図 (1)

[图5]

| | | | | | | | 7 | | | | | | | | | |
|-----------|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | child | 2 | 8 | * | 9 | L | Ħ | 14 | 16 | 01 | 12 | 13 | 15 | 3 | 6 | 2 |
| | Pio | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| | tord | 0 | 0 | 0 | .1 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 0 | 1 | 0 |
| リンクテーブル | labelid | 9 | 8 | ۴ | 9 | 7 | 5 | 8 | 8 | 4 | 9 | 7 | 7 | 2 | 2 | 1 |
| 2 | ă | 4 | 7 | 3 | 3 | 3 | 01 | 13 | 15 | 6 | 6 | 6 | 6 | 2 | 2 | 1 |
| | | | | | | | | | | | - | | | | | |
| | | | | | | | | | | | | | | | | |
| アーブル | pathid | 1 | 2 | 8 | 4 | 2 | 9 | 1 | 2 | ၈ | 4 | 5 | 4 | ω | ω | 7 |
| 中間ノードゲーング | docid. pathid | 1 1 | 2 2 | 1 3 | 1 4 | 1 5 | 1 6 | 1 1 | 1 2 | 1 3 | 1 4 | 1 5 | 1 4 | 1 5 | 1 8 | 1 7 |

"Date" "Addison-Wesley"

(I)

"Addison-Weeley"

An Introductory to Database System"

etii/

[図8]

本発明の英施例のテーブル構成の一例を示す図 (2) 無ノードアーング

| Þį | arder | wine |
|----|-------|---|
| 9 | 0 | Addison-Vesley |
| 9 | 0 | An Introductory to Database System |
| 8 | 0 | Date |
| 11 | D | Addison-Tesley |
| 12 | 0 | Foundation for Object/Relational Database |
| 14 | 0 | Date (|
| 91 | ٥ | Ватеп |
| ĺ | | |

異なってアーング

| | | , <u>.</u> | |
|----------|------|------------|--|
| attvalue | 1995 | 1998 | |
| Imbelid | 3 | 3 | |
| 32 | 8 | 6 | |

| ı, | 크 | <u> </u> | <u> </u> | L_ | | | <u> </u> | L | L |
|---------|--------|--------------------------|---------------------|-----------------|---------------------------|-----------------|----------|-----|---|
| | | | | _ | r | | · | | |
| パスロデーブル | path | bit book publisher, name | bit, book publisher | bit, book title | bih book author, lastname | bih book author | pih book | qiq | / |
| ž | pathid | 1 | . 2 | 3 | ş | 5 | 6 | 7 | 8 |
| | | | | | | | | | |

ラベルロテーブル labelid label

| qṛq | book | year | zeqs;[qnd | Sured | title | a uthor | lastrone |
|-----|------|------|-----------|-------|-------|---------|----------|
| 1 | 2 | 3 | ħ | 2 | 9 | 7 | 8 |
| | | | | | | | |

(14)

[8]

XMLデータの一例を示す図

«IELEMENT book (author», tile, publisher)»

«IATTLIST book year CDATA»

«IELEMENT artide (author», tile, year?, (shortversion i longversion))»

«IATTLIST artide type CDATA»

«IELEMENT publisher (name, address?)»

«IELEMENT suthor (giranme, address?)»

[XMLデータ]

Abook year-1985'>
- diles An Introduciony to Dalabase System «Villasauthors - dastrames Date «firstnames «Author»
cpublishers - qaamas - Addson-Westay «Inames «publisher» - «Addson-Westay «Inames «publisher»

cbook year*1986*>
«White Foundation for Object/Relational Database «Mileseauthors - databares» Date «flasharne» «duthorseauthors - databares» Dates «flasharne» «buthorepublisher» criente» Dates «Massicame» «futhor«publisher» criente» Addison-Westey «hames «publisher»
«hoot»

[國]

図8のXMLゲータをデーブルに格納した様子を示す図

もののドのデーブル

| Ω | title | suthori suthori | suthor! | 1 |
|---|---|-----------------|---------|---|
| - | An Introductory to Database System | | Date | Ä |
| 7 | Foundation for Object/Relational Database | | og C | l |
| - | | <u></u> | | l |

| | | | | ; |
|------|----------------------|-------------------|---------------------|--------------------|
| 8661 | | Addison-Westey | Darwen | |
| 5661 | | Addison-Westey | | |
| ycar | publisher address | publisher name | author2 lastname | firstname lastname |

レロントページの挑船

Fターム(参考) 5B075 ND36 PP23 QR00 QT06

(72) 発明者 石川 时种条川県川崎市中原区上小田中4丁目1番1号 含土道株式会社内